# Protein–Protein Interactions of HIV-1 Reverse Transcriptase: Implication of Central and C-Terminal Regions in Subunit Binding†

S. Patricia Becerra,‡ Amalendra Kumar,‡ Marc S. Lewis,§ Steven G. Widen,‡ John Abbotts,‡ Essam M. Karawya,‡ Stephen H. Hughes,∥ Joseph Shiloach,⊥ and Samuel H. Wilson*,‡

## Appendix: Ultracentrifugal Analysis of a Mixed Association

Marc S. Lewis§

*Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Biomedical Engineering and Instrumentation Program, National Center for Research Resources, National Institutes of Health, Bethesda, Maryland 20892, BRI-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21701, and Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892*

*Received July 17, 1991; Revised Manuscript Received September 10, 1991*

ABSTRACT: Human immunodeficiency virus 1 reverse transcriptase (RT) purified from virions is composed of a ~51 000 $M_r$ polypeptide and a ~66 000 $M_r$ polypeptide that are thought to be in heterodimer structure (Chandra et al., 1986; Hansen et al., 1988; Starnes & Cheng, 1989) and are identical except for a 15 000 $M_r$ C-terminal truncation in the smaller species (Di Marzo-Veronese et al., 1986). We prepared individual bacterial-recombinant RTs as the ~66 000 $M_r$ polypeptide (p66) or as the ~51 000 $M_r$ polypeptide (p51) and then conducted various in vitro protein–protein binding experiments. Analytical ultracentrifugation studies in 0.25 M NaCl at pH 6.5 revealed that p66 was in monomer–dimer equilibrium with $K_A$ of 5.1 × $10^4$ $M^{-1}$. p51 failed to dimerize and behaved as a monomer under these conditions. Mixing of the p66 and p51 polypeptides resulted in a 1:1 heterodimer with $K_A$ of 4.9 × $10^5$ $M^{-1}$. These results on formation of the p66/p66 homodimer and p66/p51 heterodimer were confirmed by gel filtration analysis using FPLC Superose-12 columns. Binding between p66 and individual p66 segment polypeptides also was observed using an immunoprecipitation assay. Binding between p51 and p66 in this assay was resistant to the presence of ~1 M NaCl, suggesting that the binding free energy has a large hydrophobic component. C-Terminal truncation of p66 to yield a 29-kDa polypeptide eliminated binding to p66, and N-terminal truncation of p66 to yield a 15-kDa peptide also eliminated binding to p66. The results indicate that purified individual RT peptides p51 and p66 are capable of binding to form a 1:1 heterodimer and suggest that the central region of p66 is required for this subunit binding; the C-terminal region (15 000 $M_r$) of p66 appears to be required also, as p51 alone did not dimerize.

The HIV-1 RT[1] is encoded by the pol gene in the HIV-1 genome and is expressed in low abundance in infected cells. RT is produced by proteolytic cleavage of a large gag–pol fusion protein (160–200 kDa), which is decoded by an unusual mechanism of ribosomal frameshifting from two different but overlapping ORFs, gag and pol (Jacks et al., 1988). Thus, expression of pol is regulated at least in part by the efficiency of the ribosomal frameshifting (Jacks et al., 1988). In addition to the RT, the pol gene contains two other functional ORF domains, those of protease and endonuclease. Two closely related polypeptides of apparent $M_r$ = 66 000 and 51 000 representing the RT and a 34 000 $M_r$ polypeptide representing the endonuclease are present in viral particles (Di Marzo-Veronese et al., 1986; Lightfoote et al., 1986). The N-terminal sequences of these peptides have been determined, permitting localization of RT within the central portion of the pol gene (Kramer et al., 1986), of endonuclease in the 3′ portion, and of protease in the 5′ portion.

The coding capacity of the RT ORF is 560 amino acids, corresponding to a protein of 64 484 Da and an apparent $M_r$ = 66 000 on denaturing gels. However, in virions RT is thought to occur as a dimer of the $M_r$ = 66 000 and 51 000 (p66 and p51, respectively) polypeptides mentioned above at a molar ratio of 1:1 (Chandra et al., 1986; Hansen et al., 1988). These two polypeptides are identical but for the fact that p66 has an extension at its C-end of 15 000 $M_r$ (Di Marzo-Veronese et al., 1986). This extension is known as the RNase H domain in RT (Hansen et al., 1988). The protein–protein interactions responsible for dimer formation between the two polypeptides are unknown, as is the significance of the RNase H domain on dimerization and/or enzymatic activity. It is not yet clear what the thermodynamic binding parameters are for the p66/p51 complex or, in fact, whether

† This work was supported in part by the National Institutes of Health Intramural AIDS Targeted Antiviral Program.

* To whom correspondence should be addressed at Bldg. 37, Room 4D25, NCI, NIH.

‡ Laboratory of Biochemistry, NCI, NIH.

§ Biomedical Engineering and Instrumentation Program, National Center for Research Resources, NIH.

∥ BRI-Basic Research Program, NCI–Frederick Cancer Research and Development Center.

⊥ Laboratory of Cellular and Developmental Biology, NIDDKD, NIH.

[1] Abbreviations: HIV-1, human immunodeficiency virus 1; kbp, kilobase-pair; mAb, monoclonal antibody; kDa, kilodalton(s); $M_r$, molecular weight; ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; RT, reverse transcriptase; SD sequence, Shine–Dalgarno sequence; SDS, sodium dodecyl sulfate.

isolated p51 and p66 are capable of a monomer–dimer equilibrium in vitro to form the 1:1 complex. Lowe et al. (1988) reported that in vitro proteolytic cleavage of a recombinant p66 preparation produces a p66/p51 mixture presumed to be a heterodimer. Restle et al. (1990) found by gel filtration experiments that a recombinant RT consisting of p66 and p51 behaved as a heterodimer, but this complex did not appear to be in equilibrium with the corresponding p66 and p51 monomers.

In the present report, we describe the study of complex formation in vitro between recombinant p66 and p51. The p66 protein was expressed from an *Escherichia coli* plasmid vector (pRC-RT) containing the precise coding sequence for the HXB2 HIV-1 RT enzyme. Truncated versions of p66, expressed from pRC-RT derivatives, were used in binding studies with p66. We found that the in vitro binding between the p66 and p51 polypeptides results in a 1:1 heterodimer and occurs with an association constant 10-fold greater than that observed for p66/p66 homodimer formation. Segments in the primary structure of p66 required for in vitro dimerization are described.

## Materials and Methods

*Buffers and Media.* All the buffers used during protein purification contained the following protease inhibitors: 1 mM PMSF, 10 mM $Na_2S_2O_4$, and 1 $\mu$g/mL pepstatin A. The lysis buffer contained 25 mM Tris-HCl, pH 8, 200 mM NaCl, 20% glycerol, 1% Triton X-100, 1 mM EDTA, and 2 mM DTT. Buffer A contained 50 mM Tris-HCl, pH 7.0, 1 mM EDTA, 1 mM DTT, 0.1% NP-40, and 10% glycerol. Buffer B contained 5 mM potassium phosphate, pH 6.5, 250 mM NaCl, 1 mM EDTA, 1 mM DTT, and 20% glycerol. NZY medium was purchased from the National Institutes of Health media unit; LB and M9 media were purchased from Quality Biologicals, Inc.

*Bacterial Strains and Plasmids.* Plasmid pRC23 (Crowl et al., 1985) containing the bacteriophage $\lambda$ $P_L$ promoter was used to construct expression vectors and was propagated in *E. coli* strain RRI (Maniatis et al., 1982) containing the low-copy number compatible plasmid pRK248*c*Its, which carries the gene for a temperature-sensitive *c*I repressor (Bernard & Helinski, 1979). *E. coli* strain DH5$\alpha$F′ (BRL) was used for propagation of M13-RT constructs. *E. coli* strain MV1190 (Bio-Rad) was used for propagation of initial plasmid constructs. *E. coli* strain CJ236 (Bio-Rad) was used to propagate M13 phage with uracil-containing DNA.

*Enzymes.* Restriction enzymes were purchased from Boehringer-Mannheim Biochemicals. Klenow fragment and T4 DNA ligase (highly concentrated) were purchased from New England Biolabs. An in vitro mutagenesis kit was purchased from Bio-Rad. Homogeneous Pol I large fragment (Pol I 1f) was a generous gift from Dr. William E. Brown (Carnegie Mellon University, Pittsburgh, PA). HIV-1 RT purified from virions used for initial rate comparisons was a generous gift of Dr. M. G. Sarngadharan (Bionetics Research Inc., Rockville, MD). T4 polynucleotide kinase and dNTPs were obtained from Pharmacia.

*Construction of Plasmids for Expression of HIV Reverse Transcriptase.* Preparation of DNA fragments by low-melting agarose resolution and purification through elutip-d were performed as described by the manufacturer (Schleicher and Schuell). Ligation reactions were conducted in 10 $\mu$L with 20 $\mu$g/mL DNA, 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 1 mM spermidine, 100 $\mu$g/mL BSA, 1 mM ATP, and 2000 units of T4 DNA ligase at 4 °C for a minimum of 16 h.

Bacterial transformations with M13-based constructs were performed as described by New England Biolabs (M13 Cloning and Sequencing System, A Laboratory Manual). Bacterial transformations with the initial constructs were performed as described (Maniatis et al., 1982). Bacterial transformations with pRC23-based constructs were performed following a modified version of Hanahan's method (Hanahan, 1983). Briefly, *E. coli* RRI (pRK248*c*Its) competent cells were prepared using transformation buffer composed of 100 mM RbCl, 45 mM $MnCl_2$, 35 mM potassium acetate, 10 mM $CaCl_2$, 5 mM $MgCl_2$, 0.5 mM LiCl, and 15% sucrose, at pH 5.8 (20 °C). An aliquot of 200 $\mu$L of competent cells was treated with 7 $\mu$L of DMSO before the addition of DNA (in <5 $\mu$L of ligase buffer). The mixture was freeze-shocked at −50 °C for 100 s. After being thawed once at room temperature, it was placed on ice for 5 min and then heat-shocked at 41 °C for 60 s. A total of 800 $\mu$L of LB (at room temperature) was added to the cells, and this culture was incubated at 32 °C without agitation for 1 h. An appropriate aliquot of the culture was plated on LB agar containing tetracycline and ampicillin and finally incubated at 32 °C to establish colonies. Intact pRC23 was used as a control to determine transformation efficiency of the component cells described above (~1000–10 000 transformants/ng of DNA).

The ORF for RT (N2130–3809) was prepared from the *Bgl*II–*Eco*RI DNA fragment of pHX2D, a subclone from the HXB2 HIV provirus (Ratner et al., 1985) which encompasses the entire RT region as well as its flanking sequences (provided by Bruno Starcich). A *Dra*I fragment (N2123–3173) containing the amino-terminal region of RT and seven extra nucleotides at its 5′ end (pAmino-RT) was subcloned into the *Sma*I site of pUC19. To add an ATG codon preceding the amino end of RT, a *Nco*I linker was placed at the *Kpn*I cleavage site of pAmino-RT. This step allowed an initiator codon to be positioned upstream from the first codon for RT (CCC, proline codon). However, nucleotide sequence determination showed that this ATG codon was 10 bases apart from the proline codon for RT. At the same time, a *Sca*I fragment (N2450–3810) that contained the RT carboxy-terminal region was subcloned into the *Sma*I site of pUC18 (pCarboxy-RT). In this case, a stop codon was provided from the multiple cloning area of pUC18. This also provided six extra nucleotides after the penultimate codon of RT (valine codon). The entire RT region was reconstituted by ligating the amino region from pAmino-RT to the carboxy region from pCarboxy-RT at the *Pvu*II cleavage site N2880 (pRT*).

In order to obtain the precise RT coding region flanked by start and stop codons, respectively, oligonucleotide-directed mutagenesis was performed to delete the extra sequences from either end of the RT coding region in M13-RT*. The mutagenesis was performed following instructions for Bio-Rad muta-gene in vitro mutagenesis kit. Synthetic oligodeoxynucleotides to be used as primers for the muta-gene reaction were prepared as before (Abbotts et al., 1988). Phages of M13 constructs were used as a source of templates necessary for the deletion mutagenesis; these were constructed as follows: a *Nco*I–*Hin*dIII DNA fragment from pRT* was subcloned into M13mp18 at its *Eco*RI and *Hin*dIII cleavage sites (M13-RT*); for this the *Nco*I and *Eco*RI sites were filled in with Klenow fragment in order to create blunt ends before ligation. In addition, a *Nco*I (blunt-ended)–*Eco*RV DNA fragment of 0.4 kbp containing the N-terminal region of RT was subcloned into M13mp18 at its *Eco*RI and *Sma*I cleavage sites (M13-amino*). This construct was used to perform oligonucleotide-directed mutagenesis of the 10 extra bases at

the 5′ end of RT, obtaining M13-amino. The COOH-terminal region of the coding sequence was similarly mutated to eliminate non-RT residues, producing M13-carboxy. The full-length RT coding sequence was reconstituted at a *Bal*I site, obtaining M13-RT. Appropriate deletions were confirmed by nucleotide sequence determination as described by New England Biolabs. Finally, the precise ORF for RT, an *Eco*RI–*Hin*dIII DNA fragment from M13-RT (1.7 kb), was ligated into pRC23 previously digested with *Eco*RI and *Hin*dIII. The ligation mixture was used to transform *E. coli* strain RRI (pRK248*c*Its), as described above. This final construct, pRC-RT, contained the precise coding sequence of HXB2 RT flanked by ATG and TAG codons under the control of the λ $P_L$ promoter.

In order to obtain an expression construct containing RT sequences of p51 (plasmid pRC-51), 402 nucleotides were deleted from the carboxy end of RT (p66) in pRC-RT. Removal of these sequences was accomplished by digesting pRC-RT with *Kpn*I, followed by removing their 3′ overhang end with T4 DNA Pol and *Xba*I, followed by filling in their ends with Klenow enzyme. The *Kpn*I site used is at N1277 on RT, and the *Xba*I site is at the junction between RT and the pRC23 sequences. The large fragment product of the reactions described, pRC-RT/*Kpn*I/*Xba*I, was self-ligated at its blunt ends to obtain pRC-51 transformants, as before. Thus, the 3′ end of the RT ORF in pRC-51 has lost 134 codons, with the last three codons at the C-terminal end (AAA-TTA-TGC) specifying $Lys^{424}$-$Leu^{425}$-Cys; the Cys codon is different from the authentic RT sequence and was created during plasmid construction. The coding region in pRC-51 is similar to the sequence of the natural p51 peptide. Earlier work on the C-terminal end residue of p51 polypeptides in recombinant p66/p51 mixture preparations has identified Glu 432 (B. A. Larder, personal communication), Phe 440 (Mizrahi et al., 1989), or Ala 446 (J. S. Seehra and J. M. McCoy, personal communication). The C-terminal end of our construct is, therefore, 7–21 residues shorter than these naturally derived versions of p51. Similarly, an expression construct for a peptide with the first 251 amino acids of RT, termed p29, was prepared from pRC-RT. This construct, pRC-29, has a deletion of 3 kbp between the two *Pvu*II sites of pRC-RT. The largest fragment obtained after digestion of pRC-RT with *Pvu*II was self-ligated at its blunt ends to obtain pRC-20 transformants, as before. The 3′ end of the RT ORF in pRC-29 has lost 309 codons from its 5′ end. The codons at its 5′ end specify $Ser^{251}$-Cys-Leu-Ala-Arg-Phe-Gly-Asp-Asp-Gly-Glu-Asn-Leu. The last 12 residues are different from the authentic RT sequence and were added during plasmid construction. The p15 construct corresponding to residues 425 to the COOH-terminal end of p66 has been described (Becerra et al., 1990).

*Expression of Reverse Transcriptase Peptides (i.e., p66, p51, p29, and p15).* Bacterial cells containing RT expression vectors were grown in NZY media containing 50 μg/mL ampicillin at 30 °C to early logarithmic phase such that $OD_{600nm}$ was only 0.1. Cultures then were incubated in a 65 °C water bath for a brief period with agitation until the culture temperature reached exactly 42 °C. Then cultures were transferred to an air-flow incubator at 42 °C with hard agitation (300–400 rpm). After incubation for 2–3 h, the cells were collected by centrifugation and washed in a solution containing 10 mM Tris, pH 7.6, 1 mM EDTA, and 100 mM NaCl (TEN). Cells were sonicated using a probe as described below. For protein analysis by SDS–PAGE, the cell pellet was suspended in TEN at 30 μL per 5 mL of culture per 0.1 $OD_{600nm}$, plus 62.5 mM Tris, pH 6.8, 10% glycerol, 2% SDS, 5% 2-mercaptoethanol, and 0.002% bromophenol blue.

For protein radiolabeling, bacterial cells containing pRC-51 or pRC-29 expression vectors were grown in M9 media to $OD_{600nm}$ = 0.1, as for pRC-RT. After the culture temperature had reached 42 °C, L-[$^{35}$S]methionine (1037 Ci/mmol, Amersham Corp.) was added at 50 μCi/mL of culture; incubation at 42 °C and cell harvesting proceeded as above. Incorporation of $^{35}$S into p51 or p29 was analyzed by immunoprecipitation with monoclonal antibodies followed by SDS–PAGE (Becerra et al., 1988). Incorporation of L-[$^{3}$H]leucine into p15 was conducted and analyzed in a similar fashion, using 100 μCi of [$^{3}$H]leucine/mL of culture (48 Ci/mmol, Amersham Corp.).

*Protein Purification.* A total of 10 g of *E. coli* cells, induced 3 h at 42 °C, was suspended in 100 mL of lysis buffer and probe-sonicated with five bursts, 30 s each, at 50 W with the tube immersed in an ice-water bath. (Alternatively, an extract of a 50-g batch of cells was prepared using a French press.) The lysate was kept at 0 °C for 15 min, and then particulate material was removed by centrifugation at 10000*g* for 10 min. The supernatant fraction was dialyzed against buffer A containing 75 mM NaCl at 4 °C for 3 h. The dialyzed extract was centrifuged at 10000*g* for 5 min, and the supernatant fraction was passed through a 0.22-μm filter (MILLEX-GV from Millipore Corp.). The filtered solution was layered on a Q-Sepharose fast flow (Pharmacia) column (20-mL bed volume), and the flow-through fraction was collected (pool I).

Pool I was layered on a ssDNA–agarose column (15-mL bed volume, Pharmacia) equilibrated with buffer A containing 75 mM NaCl. The column was washed with 1 column volume of buffer A containing 75 mM NaCl and then with 1 column volume of buffer A containing 100 mM NaCl. Finally, p66 was eluted with a linear gradient of 150–600 mM NaCl in buffer A (total volume 100 mL). Fractions of 2.5 mL were collected at a flow rate of 0.75 mL/min.

Several fractions containing the peak (but not trailing edge) of p66, as observed on Coomassie blue stained polyacrylamide gels, were pooled (pool II, 30 mL). Pool II was dialyzed against buffer A containing 75 mM NaCl, and ammonium sulfate precipitation was then performed. Polypeptide p66 appeared in a precipitated fraction at 35–65% saturation, and the precipitated protein was dissolved in buffer A plus 75 mM NaCl and layered on a Superose-12 (Pharmacia) column (25-mL bed volume) equilibrated with buffer A containing 75 mM NaCl and running at 0.35 mL/min. p66-containing fractions from the Superose-12 column (pool III) were pooled, dialyzed against buffer B, and layered on a hydroxylapatite column equilibrated with buffer B. The column was washed with buffer B, and elution of p66 was accomplished with a gradient of 5 mM (pH 6.5)–100 mM (pH 7.0) potassium phosphate in buffer B at 0.6 mL/min. Fractions were of 0.7 mL, and the peak fractions with p66 alone were pooled and stored at −70 °C (pool IV). Recovery in pool IV was 0.5 mg/g of packed cells, representing approximately 8% yield from the crude soluble extract. A side fraction with p66 and a 51 000 $M_r$ species was collected also and was obtained in about the same yield as pool IV.

p51 was purified by the same procedure from 20 g of cells induced to express p51 from pRC-51.

*Immunoblot Analysis.* Proteins were transferred from polyacrylamide gels onto nitrocellulose paper by the method of Towbin et al. (1979); the blotted paper was treated as described (Swack et al., 1985) and reacted with a monoclonal antibody against HIV RT, mAb (RT) (New England Nuclear,

Inc., Boston, MA), at 1:2000 dilution for 1 h. The paper was incubated with biotinylated anti-mouse IgG (Cappel Laboratories) at a 1:500 dilution for 1 h and then with an avidin-biotinylated horseradish peroxide complex (Vectastain ABC Kit, Vector Laboratories) for 1 h. Color was developed by reaction with 0.3% HRP color development reagent (Bio-Rad) and hydrogen peroxide. Purified rabbit IgG against Pol I was generously provided by Dr. William E. Brown (Carnegie Mellon University).

*Complex Formation by RT Polypeptides and Analysis of Products by Immunoprecipitation.* Purified RT (p66) was mixed with bacterial protein extract containing $^{35}$S-p51 at an approximate ratio of 1:1, 0.8 $\mu$M each. The incubation for complex formation was performed at 30 °C for 4 h, in a final volume of (15 $\mu$L), containing 5 $\mu$L of total soluble extract from $^{35}$S-labeled cells. Each mixture contained the components of lysis buffer, in addition to proteins as indicated, and 3.5 × $10^5$ dpm $^{35}$S-labeled proteins. The antibody mAb 50 (Ferris et al., 1990) toward the COOH-terminal end of p66 was added (5 $\mu$L), and the immunoreaction was allowed to proceed at 4 °C for 16 h. Immunoprecipitation using Staph A–Sepharose beads (Sigma) was performed as described (Becerra et al., 1988). Immunoprecipitated proteins were washed, then dissolved in SDS sample buffer, and resolved by SDS–PAGE. Autoradiography generally was for four days. Other monoclonal antibodies, mAb 19 (Ferris et al., 1990) or mAb (RT), were used as appropriate. The concentration of p51 or p29 in the soluble extract was obtained from the relative amounts of these proteins in the bacterial protein extracts as determined by Coomassie blue staining after SDS–PAGE.

*Processivity Analysis with an M13 DNA Template.* M13mp2 DNA (viral plus strand) and primer 115 were generous gifts from Dr. Thomas A. Kunkel (National Institutes of Environmental Health Sciences, Research Triangle Park, NC). A synthetic DNA primer, complementary to sequence positions 129–115 (primer 115) of the M13mp2 DNA template, with its 3′-hydroxyl group corresponding to position 115, was 5′-end-labeled using [$\gamma$-$^{32}$P]ATP (ICN Radiochemicals) by the procedure of Maxam and Gilbert (1980). DNA was extracted with phenol, and the residual phenol was removed with ether; the DNA was precipitated with ethanol. In vitro DNA synthesis on M13 DNA and analysis of products by gel electrophoresis in a 12% polyacrylamide/7 M urea gel were performed as described by Bebenek et al. (1989).

*Analytical Gel Filtration.* Purified p66 and p51 were used for subunit binding by gel filtration. Unless otherwise indicated, 35 $\mu$g of p66 or 35 $\mu$g of p51 or a mixture containing 20 $\mu$g of p66 and 20 $\mu$g of p51 was adjusted to 100 $\mu$L in 100 mM potassium phosphate buffer, pH 7.5/200 mM NaCl/0.1 mM EDTA/0.1 mM DTT/5% glycerol and was incubated at 30 °C for 2 h. These samples were used for analytical gel filtration. Gel filtration was performed using a Superose-12 column on an FPLC system (Pharmacia). The column was equilibrated and eluted with the above buffer at 25 °C. The flow rate was 0.4 mL/min, and 0.5-mL fractions were collected. This column was routinely calibrated using three standard proteins: IgG, BSA, and ovalbumin.

*Analytical Ultracentrifugation.* Analytical ultracentrifugation was performed using a Beckman Model E analytical ultracentrifuge equipped with a scanning absorption optical system. Data were acquired from the scanner output using a Metrabyte DAS-8 12-bit analog to digital converter in a 10-MHz AST Premium/286 computer as an acquisition system. Using the rapid scan rate, 90 000 data points were acquired in the 18 s required to scan from the outer reference hole to the inner reference hole of the counterbalance. Each recorded point was the average of 100 acquired points; the actual data density was 425 points/cm of radial distance in the cell. The inner and outer reference points and the region of interest were selected in the editing of the scan, and the data were saved in the form of millivolts as a function of radial position.

At the time of analysis, the data were converted to optical density as a function of radial position using a conversion factor obtained in calibration experiments. Further data manipulation and data analysis by mathematical modeling were performed using MLAB operating on the data acquisition computer. MLAB for the PC is a new and enhanced version of the MLAB system which was originally developed to operate on DEC-10 and DEC-20 mainframe computers (Knott, 1979). It is powerful and flexible, relatively fast for an interpretive system, and well suited for ultracentrifugal data analysis in which two or more sets of data must be simultaneously fit, with weights, to different mathematical models.

The experiments were run using a six-hole rotor with five double-sector cells and a counterbalance at a rotor speed of 12 000 rpm. p51 with an absorbance of approximately 0.2 at 280 nm was loaded in one cell; p66 and approximately equimolar mixtures of p51 and p66 with absorbancies of approximately 0.2 and 0.1 $A_{280nm}$ were loaded in the other cells. Two concentrations were used for p66 and for the mixture. This was done because it has been demonstrated that the simultaneous fitting of data from experiments involving multiple loading concentrations—requiring the simultaneously fit models to have an equilibrium constant(s) as a global fitting parameter(s) and cell reference concentrations and base-line error terms as local fitting parameters—represents the most stringent criterion possible for establishing the validity of what appears to be a reversible associating system (Roark, 1976).

The time required for the attainment of equilibrium was established by running at a given rotor speed until the scans were invariant for 24 h; this was achieved after 74 h for solution column lengths of approximately 4 mm. Derivations of the mathematical models used for fitting and pertinent experimental details are described in the Appendix. Sedimentation was at 5 °C in a solution containing 5 mM potassium phosphate buffer, pH 6.5, 250 mM NaCl, 1 mM EDTA, 1 mM DTT, and 5% glycerol.

*Other Methods.* Protein concentration was determined by the method of Lowry et al. (1951). SDS–PAGE of proteins was performed according to Laemmli (1970). Protein markers were from Bio-Rad. Activity gel analysis was exactly as described (Karawya et al., 1983) using 10% SDS–polyacrylamide gels containing activated calf thymus DNA as template-primer and [$^{32}$P]dNTP; $\beta$-polymerase (Kumar et al., 1990), HIV RT from MicroGeneSys, and Pol I 1f (Klenow enzyme) were used as standards. Initial rates of DNA polymerase activity were measured as described by Majumdar et al. (1989) with poly(rA)·oligo(dT) as template-primer. Extinction coefficients were obtained from amino acid analysis and spectral analysis of purified p66 and p51 and were found to be 1.36 and 0.93 $A_{280nm}$, respectively, for 1 mg/mL solutions. N-Terminal sequencing of a liquid sample of p66 was conducted in an Applied Biosystems, Inc. 477A instrument using a portion of pool IV.

## Results

*Expression of HIV-1 RT in E. coli.* Before the HIV-1 RT coding sequence (HXB2) was inserted into pRC23, the ends were manipulated so that initiation and termination codons

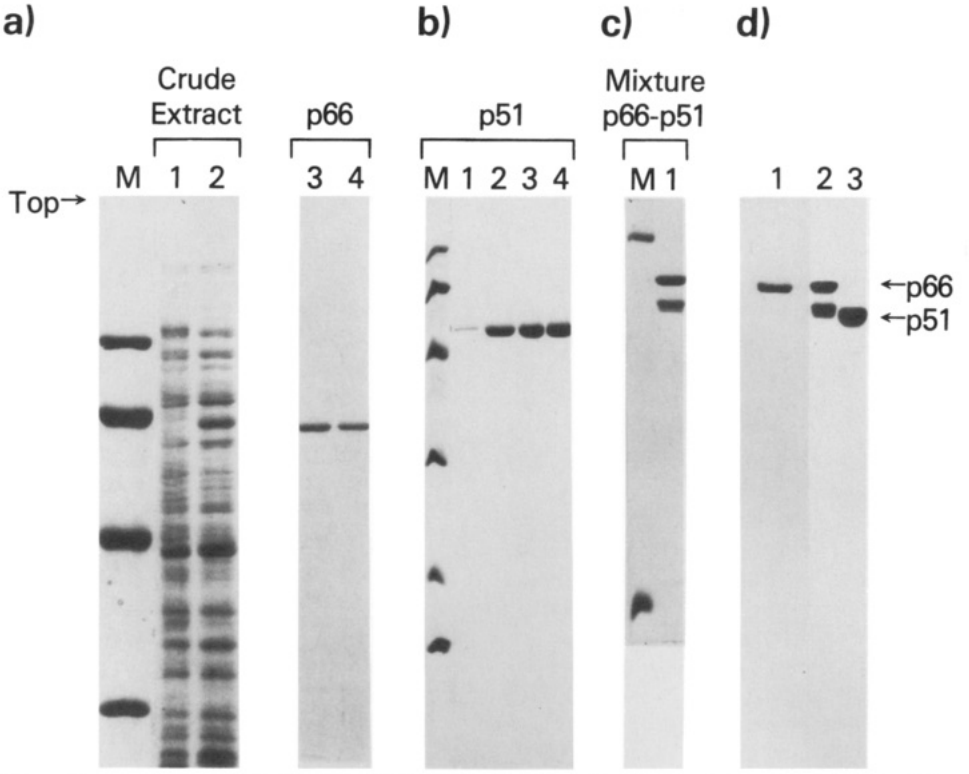



FIGURE 1: SDS–PAGE and immunoblot analysis of purified HIV-1 RT p66, p51, and p66/p51 mixture. Photographs of Coomassie blue stained gels are shown. Panel a: Lanes 1 and 2 show analysis of crude *E. coli* RRI (pRK248*c*Its; pRC-RT) cell extracts either noninduced (30 °C) or induced (42 °C) for expression of RT; extract from the same amount of cell pellet was added to each lane. Lanes 3 and 4 show purified p66 (pool IV) at two levels of protein: 3 and 1 μg, respectively. Panel b: Lanes 1–4 represent 0.1, 1, 2, and 3 μg, respectively, of purified p51. Panel c: Lane 1 is 5 μg of the purified p66/p51 mixture. In panel d, a photograph of immunoblot with a 1:2000 dilution of monoclonal antibody to HIV RT mAb (RT) as probe with 4-chloro-1-naphthol staining is shown. Lane 1, p66; lane 2, p66/p51 mixture; lane 3, p51. Lane M contains molecular mass markers as follows: phosphorylase *b* 97.4 kDa; bovine serum albumin, 66.2 kDa; ovalbumin, 42.7 kDa; carbonic anhydrase, 31.0 kDa; soybean trypsin inhibitor, 21.5 kDa; lysozyme, 14.4 kDa.

were placed in-frame upstream and downstream, respectively, from the precise RT coding sequence. The final expression construct, pRC-RT, had the RT ORF under the control of the λ $P_L$ promoter. Deletion constructs derived from pRC-RT were used for production of two polypeptides corresponding to segments of p66 RT as follows: N-terminal, 51 kDa; N-terminal, 29 kDa. Our purification protocol for the recombinant RT yielded purified p66 and also a side fraction of p66 along with a lower $M_r$ polypeptide of 51 000. A typical extract used for purification of p66 and the final fractions of p66, p51, and the p66/p51 mixture were analyzed in the experiments shown in Figure 1. The p66 RT polypeptide in the crude extract and purified p66 had the same apparent molecular weight, and as expected, each of the p66 and p51 peptides in the RT preparations immunoreacted with a monoclonal antibody to HIV-1 RT, mAb (RT) (Figure 1). p51 in the p66/p51 mixture probably arose from proteolysis of p66 in the *E. coli* extract (Lowe et al., 1988).

We selected the purified p66 RT for further characterization. Sequence analysis of this purified protein revealed that the $NH_2$-terminal residue was Pro and that the sequence of the first 14 residues precisely matched the deduced sequence for the RT coding region from HXB2 proviral DNA, PISPIETVPVKLKP (Lightfoot et al., 1986); no other sequences were detected in this analysis. The rate of enzymatic activity of p66 with poly(rA)-oligo(dT) as template-primer was about 800 nmol of dTMP incorporated $min^{-1}$ (mg of protein)$^{-1}$ at 37 °C, which is similar to that of the enzyme purified from virions (Majumdar et al., 1989). Activity gel analysis (not shown) revealed that p66 contained one active DNA polymerase polypeptide, i.e., $M_r \cong 66\,000$, indicating that Pol I was not a contaminant in the preparation. Processivity analysis of p66 is shown in Figure 2. Experiments were conducted using an M13 DNA template, and p66 was compared with HIV-1 RT purified from virions. p66 appeared to be moderately more processive than the virion enzyme, as p66 showed less termination at early positions and an increased amount of longer product molecules. p66 exhibited a termination pattern similar to that of the virion enzyme. Yet, minor differences could be seen in the termination pattern for the two enzymes (Figure 2, lanes 1–4). For example, the virion enzyme showed strong termination sites at positions 89–92, while p66 showed strongest sites at positions 89 and 91; the virion enzyme showed a strong site at position 70, but termination by p66 was less pronounced at this position. The basis for these modest differences between p66 and the virion RT is not known, but the differences were not due to contamination by Pol I, which gives a different termination pattern on this template (Figure 2, lanes 5 and 6).

*Protein–Protein Binding Properties of p66 and p51.* Protein–protein binding properties of the recombinant RT peptides were examined by analytical sedimentation analysis in a Model E ultracentrifuge and by gel filtration analysis in an FPLC Superose-12 column. The standard errors of the concentrations ($SE_c$) in the ultracentrifugal analysis were determined in the scanner calibration experiments and are described by the equation $SE_c = 0.003 + 0.005c$, and the values in the weighting vector ($W_c$) for each fit were calculated from the normalized reciprocals of the variances using the equation $W_c = 9/(9 + 15c + 25c^2)$ where $c$ is the concentration in absorbency units at 280 nm. Figures 3–5 illustrate such weighted fits, which are described in detail in the Appendix.

Figure 3 shows the fit of the data for p51 using eq A1 and demonstrates that it is a homogeneous monomer. The actual

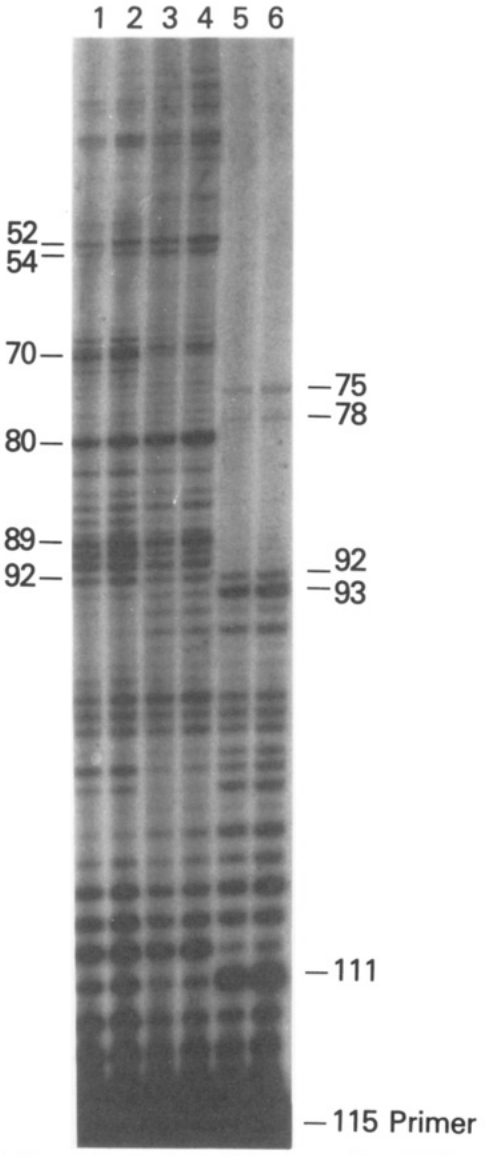



FIGURE 2: Processivity analysis of RT preparations. DNA synthesis reactions were carried out with labeled primer 115, and the products of processive synthesis were examined by gel electrophoresis as described under Materials and Methods. An autoradiogram is shown. DNA synthesis reactions were conducted with the following enzymes and incubation times: lanes 1 and 2, HIV-1 reverse transcriptase purified from virions, 10 and 30 min; lanes 3 and 4, p66 pool IV, 10 and 30 min; lanes 5 and 6, *E. coli* Pol I 1f, 10 and 30 min.

fit is shown in the lower section, and a plot of the distribution of the residuals is shown in the upper section. Fitting for the value of $A$ and using a molecular mass of 49663.1 daltons from the amino acid sequence and the experimentally determined value of 1.0278 g/cm for the buffer density at 5 °C gives a value of 0.754 $cm^3/g$ for the partial specific volume at that temperature. This differs significantly from the calculated compositional partial specific volume of 0.718 $cm^3/g$ at 5 °C. The reason for this difference is not readily apparent, particularly when the calculated partial specific volume of p66 proved to be an optimal value. The experimentally determined value was used in the calculations involving p51.

Panels A and B of Figure 4 show the fit of the data for p66 using eq A3 with a monomer molecular mass of 64501.2 daltons, a calculated compositional partial specific volume of 0.717 $cm^3/g$, and the experimentally determined buffer density at 5 °C. Figure 4A is for a loading concentration of ~0.2 absorbance units; Figure 4B is for half that concentration. The simultaneous fitting procedure demonstrated that we observed

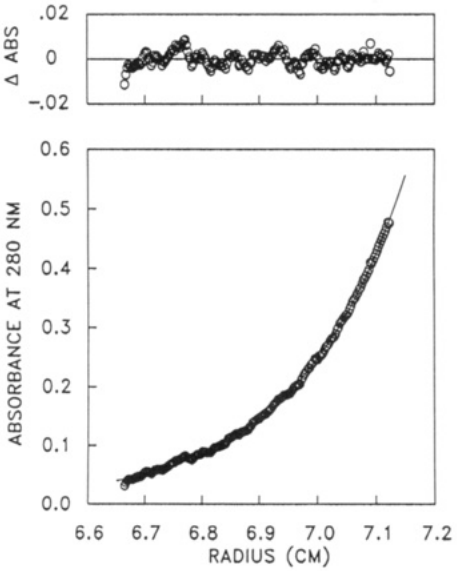



FIGURE 3: Analytical untracentrifugal analysis of p51. The analysis was performed in a Beckman Model E ultracentrifuge operated at 5 °C. The loading protein concentration was 0.29 $A_{280nm}$ (p51). Concentration distribution (lower graph) and distribution of the residuals (upper graph) as functions of radial position for p51 at ultracentrifugal equilibrium after 74 h at 12000 rpm at 5 °C are shown. The fitting line is for a thermodynamically ideal monomer.

Table I: Summary of Binding Parameters of Recombinant HIV-1 Reverse Transcriptase Peptides

| reaction parameter[a] | 2p66 ⇌ $p66_2$ | p51 + p66 ⇌ p51/p66 |
|---|---|---|
| ln $K_a$ | 10.83 ± 0.02 | 13.11 ± 0.07 |
| $K_A$ | $5.06 \times 10^4$ $M^{-1}$ | $4.92 \times 10^5$ $M^{-1}$ |
| $\Delta G°_{5°}$ | −5.99 ± 0.01 kcal $mol^{-1}$ | −7.24 ± 0.04 kcal $mol^{-1}$ |

[a] Parameters were obtained using data fits to models as described in the Appendix. The standard state was taken to be 1 mol/L of reactant monomer.

a reversible monomer–dimer equilibrium. Additionally, attempts to fit with equations involving species larger than the dimer indicated that such species were not present. Attempts to obtain an experimentally determined value for the partial specific volume indicated that the calculated compositional value was optimal. The values of the equilibrium constant and free energy change are given in Table I. It is noted that fits in Figure 4A,B show more deviation at radial positions greater than 7.1 cm. Our experience has demonstrated that deviation in this region is caused by an artifact introduced by the window holders in the rotor. The fact that the deviation occurs at the greatest concentrations also means that these data points have the least weight, and the effect of the deviation upon the parameter values is minimal.

Panels A and B of Figure 5 show the data fit for the p51/p66 interaction using eq A22 and indicate that p51 and p66 were in monomer–dimer equilibrium. As before, the initial concentrations for Figure 5B were half those for Figure 5A. A weight-average partial specific volume of 0.733 $cm^3/g$ for the p51/p66 complex was calculated as described in the Appendix. Values of $4.62 \times 10^4$ and $8.80 \times 10^4$ were used for the molar extinction coefficients of p51 and p66, respectively, for the calculation of the value of $D_{AB}$; the values of $c_{0,A}$ and $c_{0,B}$, used in this term, were calculated using eqs A23 and A24 with appropriate adjustments for dilution. The value of ln $k_{12}$ was that obtained from fitting the p66 data. The value of ln $k_{AB}$ was a global fitting parameter, and the values of $c_{b,B}$ and $\epsilon$ were local fitting parameters. The quality of the joint fits demonstrated that the monomer–dimer fitting model was

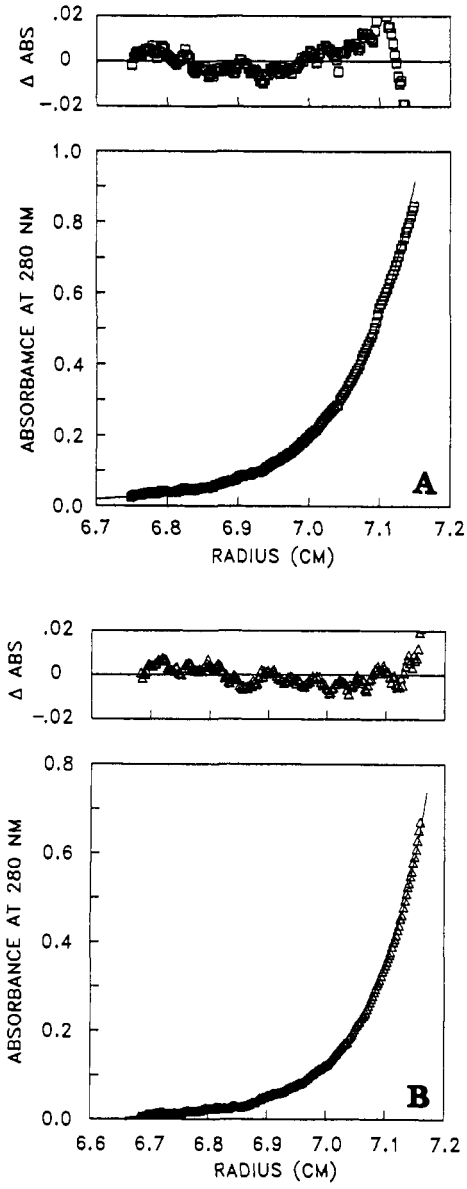



FIGURE 4: Panel A: Concentration distribution and distribution of the residual as functions of radial position for p66 at ultracentrifugal and chemical equilibrium after 74 h at 12000 rpm at 5 °C. The fitting line is for a thermodynamically ideal homogeneous monomer–dimer association of p66 and is a simultaneous fit with the data shown in panel B. The loading protein concentration was 0.29 $A_{280nm}$. Panel B: The same as panel A except at half the loading concentration.

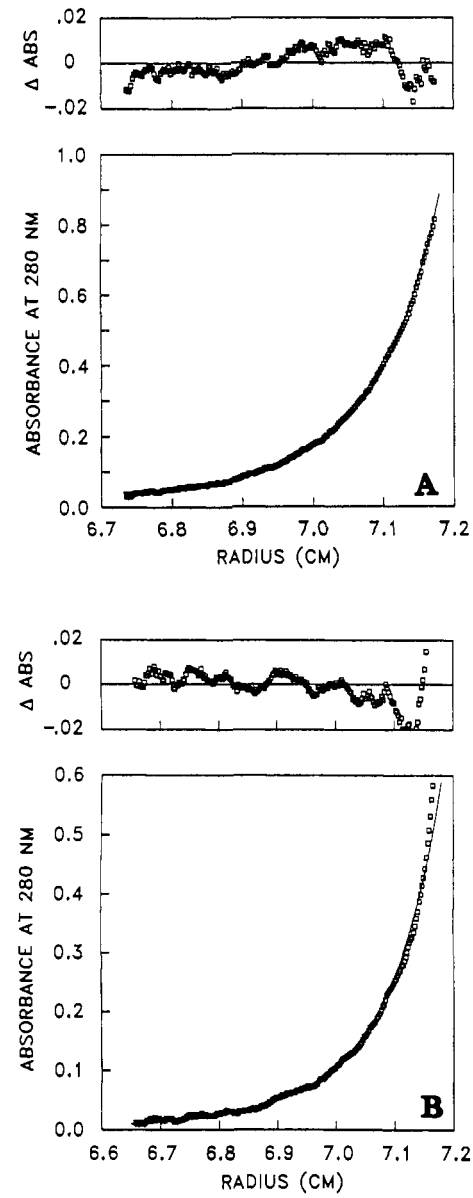



FIGURE 5: Panel A: Concentration distribution and distribution of the residuals as functions of radial position for an approximately equimolar mixture of p51 and p66 at ultracentrifugal and chemical equilibrium after 74 h at 12000 rpm at 5 °C. The fitting line is for a thermodynamically ideal heterogeneous association of p51 combined with a homogeneous monomer–dimer association of p66 and is a simultaneous fit with the data shown in panel B. Panel B: The same as panel A except at half the loading concentration.

appropriate. Figure 5B shows some of the same distortion described above. Fitting each data set with eq A25 and calculating the values of the sum of the initial concentrations of p51 and p66 gave values that were in nominal agreement with the sums of the individual values obtained by the use of eqs A23 and A24, thus indicating that mass was conserved and that the constrained model was appropriate. The values of equilibrium constant and free energy change (Table I) indicate that p51/p66 heterodimerization has a greater free energy change than p66 homodimerization by 1.25 kcal $mol^{-1}$.

Gel filtration analysis using an FPLC Superose-12 column provided resolution in the size range of ~40000 $M_r$ to 150000 $M_r$ (Figure 6a). Analysis of a p66 sample at ~2 mg/mL using Superose-12 FPLC chromatography (Figure 6b) revealed that most of the protein was in a $M_r \approx 130000$ complex corresponding to the homodimer; only a small amount (~10%) of protein was in the monomer position. At lower sample protein concentrations, less $M_r = 130000$ complex was observed, and at 0.6 mg/mL most of the p66 corresponded to a monomer. Gel filtration analysis of the p51 preparation at 2 mg/mL revealed that p51 behaved as a monomer (Figure 6c). Mixing of p51 with p66 at 1 mg/mL each resulted in formation of some heterodimer complex, since a portion of the p51 now was recovered in the gel filtration fraction (no. 17) corresponding to the dimer (Figure 6c inset). These results directly demonstrating complex formation between p66 and p51 are consistent with and confirm the results of our ultracentrifugation analysis, which depend upon mathematical modeling procedures.

*In Vitro Binding between p66 and Segment Polypeptides.* To further examine the interaction between p66 and p51, we prepared $^{35}S$-labeled p51, as described under Materials and Methods. We also prepared an $^{35}S$-labeled ~31000 $M_r$ peptide, termed p29, representing further C-terminal truncation of p51 (see Figure 7). As expected, the $^{35}S$-labeled p51

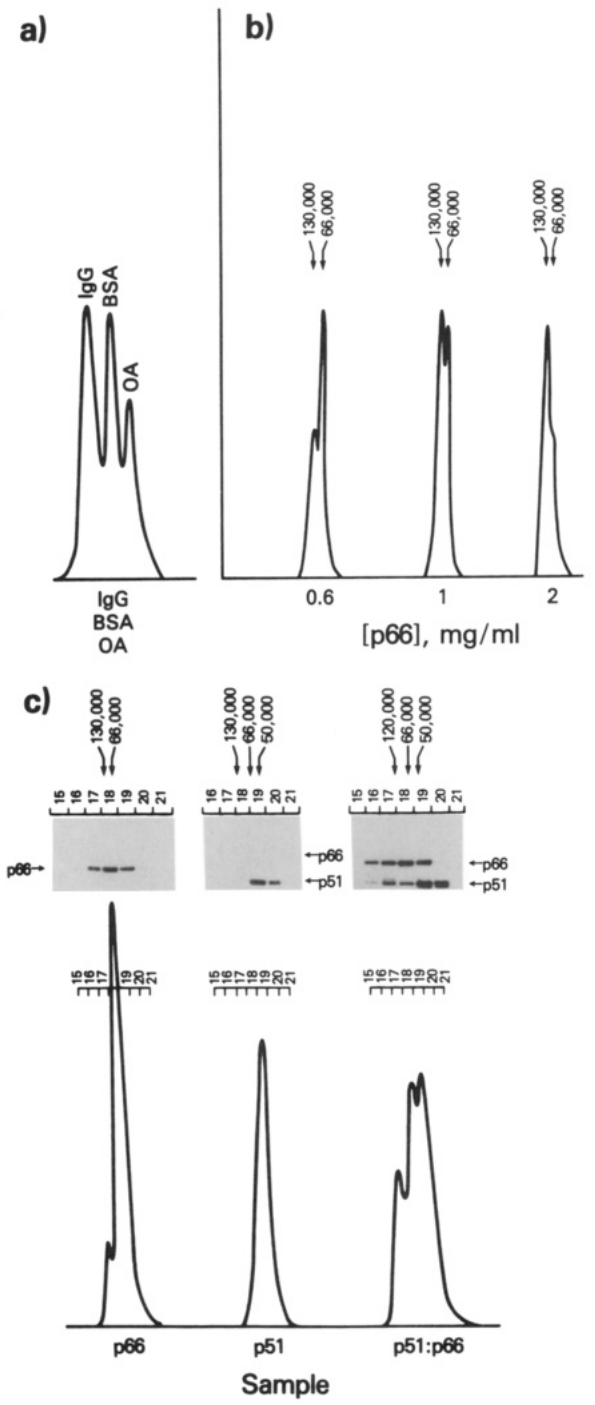



FIGURE 6: A composite showing analysis for binding by gel filtration using a Superose-12 column on an FPLC system. Chart recordings of absorbance (280 nm) for seven individual chromatographic experiments are presented, and photographs of immunoblotting analysis are shown. (a) Chromatographic pattern of a 100-$\mu$L mixture of three marker proteins: 25 $\mu$g of IgG (145 000 $M_r$); 25 $\mu$g of BSA (66 000 $M_r$); and 20 $\mu$g of OA (45 000 $M_r$). Full-scale absorbance was 0.01 $A_{280nm}$. (b) Chromatographic pattern of p66. Column conditions were as described below. A total of 100 $\mu$L of three different concentrations of p66 was layered on the column as indicated at the bottom of the panel, and full-scale absorbance was 0.1, 0.1, and 0.2 $A_{280nm}$, respectively. The sizes indicated at the top were taken from a calibration of the column with marker proteins. (c) Chromatographic pattern of the p66 preparation or the p51 preparation or a mixture of the two preparations. Full-scale absorbance was 0.01$A_{280}$. Samples of p66, p51, or these mixed (p66/p51 as described under Material and Methods) were applied to the column, and 0.5-mL fractions were collected. A total of 200 $\mu$L of each fraction was analyzed on 12.5% polyacrylamide–SDS gels, and immunoblots were prepared using mAb (RT); column fractions analyzed are shown, as are the p51 and p66 polypeptides. Sizes indicated at the top are as in panel b.

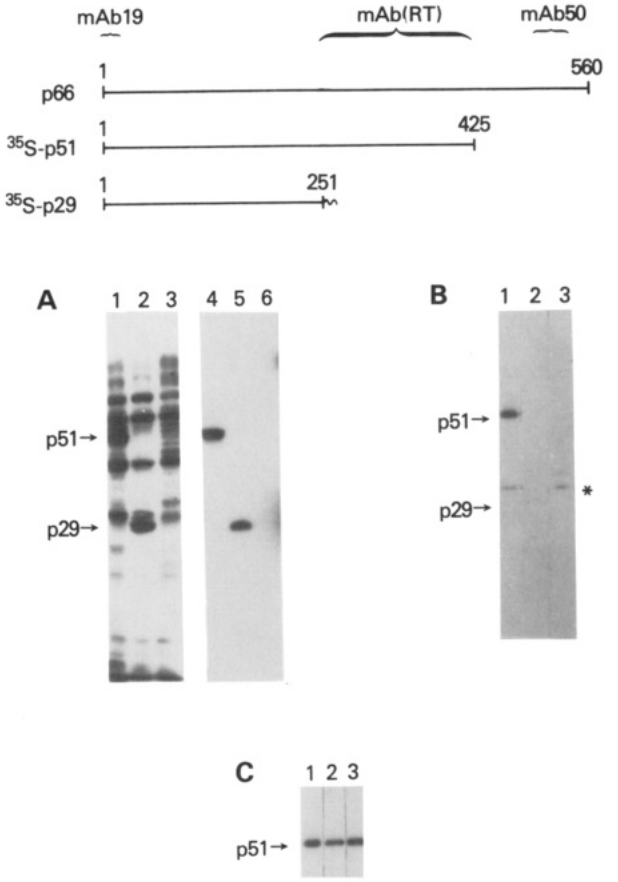



FIGURE 7: In vitro complex formation between purified p66 and p51 revealed by immunoprecipitation. The structures of p66, p51, and p29 proteins encoded by the constructs used in this assay are given. Numbers indicate amino acids. The brackets at the top show the region in RT representing the epitopes of monoclonal antibodies mAb 19 and mAb 50 as described (Ferris et al., 1990) and of mAb (RT) (NEN). Although the epitope(s) recognized by mAb (RT) has not been described before, RT deletion constructs containing amino acid residues 1–252 and 428–560 produced proteins that did not immunoreact with mAb (RT), indicating that its epitope falls between residues 252 and 428 (unpublished results). Autoradiograms of SDS–PAGE are shown. (a) Expression of recombinant $^{35}$S-labeled p51 and $^{35}$S-labeled p29 proteins in *E. coli*. Cultures of RRI (pRK248*c*Its) containing pRC-51 (lane 1), pRC-29 (lane 2), and pRC23 (lane 3) were grown and induced for expression in the presence of [$^{35}$S]Met as indicated in the text. Soluble protein extracts obtained after sonication were subjected to 10% SDS–PAGE followed by fluorography. $^{35}$S-Labeled proteins from RRI (pRC-51) (lane 4), RRI (pRC-29) (lane 5), and RRI (pRC23) (lane 6) were immunoprecipitated using N-terminal mAb19 and applied to the same gel. The arrow indicates p51 and p29 protein positions. (b) In vitro complex formation in the presence of 0.2 M NaCl. $^{35}$S-labeled p51 in the total soluble extract was incubated with or without purified p66 (pool IV) to allow complex formation, as described in the text. The proteins were then immunoprecipitated using the C-terminal mAb 50 and analyzed on 10% SDS–PAGE followed by fluorography. The weak band (indicated by asterisk) in all lanes is due to a small amount of unspecific, non-RT-precipitable material (unpublished results). Lane 1, $^{35}$S-labeled protein extract of RRI (pRC-51) and p66 protein; lane 2, $^{35}$S-labeled protein extract of RRI (pRC-29) and p66 protein; and lane 3, $^{35}$S-labeled protein extract of RRI (pRC23) and p66 protein. The arrow indicates the p51 protein position. (c) Effect of NaCl on the in vitro complex formation. Increasing concentrations of NaCl were used during the incubation of p66 and RRI (pRC-51) soluble $^{35}$S-labeled protein extract to allow the formation of the p66–p51 protein complex. The proteins were then immunoprecipitated and analyzed as in panel b. Lane 1, 0.375 M NaCl; lane 2, 0.75 M NaCl; and lane 3, 1.125 M NaCl. The arrow indicates the $^{35}$S-labeled p51 position.

and $^{35}$S-labeled p29 peptides could be immunoprecipitated from their respective soluble crude extracts by the monoclonal antibody with an epitope in the N-terminal region, mAb 19

(Figure 7a, lanes 4 and 5), whereas control (pRC23) crude extract did not yield any immunoprecipitated proteins (Figure 7a, lane 6). For binding experiments, we added purified p66 to the soluble crude extract containing $^{35}$S-labeled proteins; p66 was added to a final concentration of 0.1 mg/mL, and the mixture was incubated for 4 h at 30 °C. Immunoprecipitation then was conducted with the monoclonal antibody (mAb 50) against an epitope in a C-terminal portion of p66 not present in p51 or p29. With this antibody, p51 could appear in the immunoprecipitate only by virtue of its binding to p66. We found labeled p51 in the immunoprecipitate with the addition of p66 to the p51 crude extract (Figure 7b, lane 1), and as expected, no labeled p51 was precipitated without the addition of p66. These results indicate that p51 formed a protein–protein complex with p66. Interestingly, when labeled p29 extracts were used, the p29 peptide was not precipitated from the crude extract by virtue of binding to p66 (Figure 7b, lane 2). Similar experiments with the C-terminal peptide p15 and mAb 19 against an N-terminal epitope failed to reveal binding to p66 (data not shown). These observations that p29 and p15 did not form a stable complex with p66, whereas p51 did, suggest that a domain required for protein–protein binding had been eliminated from the p29 and p15 peptides. Interactions involved in p51 to p66 binding were evaluated by increasing the concentration of NaCl in the binding mixture. Binding was not reduced even by the presence of 1 M NaCl (Figure 7c). Since NaCl at this concentration is considered to neutralize most of the ionic interactions potentially involved, these results suggest that the p66/p51 interaction has a large hydrophobic component.

*DNA Polymerase Activity and p51–p66 Binding.* To examine the functional significance of p66 and p51 binding, the two purified proteins were mixed and incubated for 1 h at 30 °C under the usual binding conditions, and DNA polymerase activity then was measured. The activity of 0.2 μg of p66 was stimulated by p51; maximal stimulation was 2–3-fold, and the amount of p51 corresponding to half-maximal stimulation was 0.15 μg. p51 alone had 20-fold less polymerase activity than p66 alone. These results suggesting that the p51/p66 complex is slightly more active than the p66/p66 complex are in accord with the results of Lowe et al. (1988); they used an in situ proteolysis approach to convert their p66/p66 preparation to p51/p66 and found that the latter preparation was 2–3-fold more active than the former.

## Discussion

The concept that the HIV-1 RT is purified from virions as a dimer of ~66-kDa and ~51-kDa polypeptides came from gel filtration analysis by Chandra et al. (1986) and from glycerol gradient centrifugation analysis by Hansen et al. (1988). In both cases, the stoichiometry of the 66-kDa to the 51-kDa peptide in the enzymatically active complex of ~120 kDa appeared to be 1:1. However, it was not clear whether this complex might be contaminated by homodimers of the two polypeptides. We have shown here, and others have recently shown (Restle et al., 1990), that recombinant versions of the 51-kDa polypeptide do not undergo homodimerization under conditions where these peptides bind tightly to the recombinant 66-kDa peptide. The implication is that the 66-kDa–51-kDa recombinant RT is indeed a 1:1 heterodimer. Our studies of protein–protein complex formation revealed that individually purified p66 and p51 are capable of binding to form a 1:1 heterodimer, as revealed by both analytical ultracentrifugation and gel filtration analyses. The association constant measured by sedimentation analysis ($4.92 \times 10^5$ M$^{-1}$) corresponded to a $-\Delta G°_{5°}$ of 7.24 kcal mol$^{-1}$.

The fact that p66/p51 binding can occur in the presence of a high concentration of NaCl (Figure 7c) points to the importance of hydrophobic interactions in the overall free energy of binding. In considering which portions of the primary structure of RT might contribute to this type of protein–protein interaction, our results indicated, first, that a ~20 000 $M_r$ C-terminal truncation of p51 eliminated all complex formation with p66 and, second, that a peptide corresponding to the C-terminal 15 000 $M_r$ of p66 was not able to bind to p66. Thus, these results suggest that the central region of RT may contain sequences required for the protein–protein interactions involved in dimerization. Yet, p51 containing this central 20-kDa region was not able to dimerize under our conditions, clearly indicating that the 15-kDa C-terminal segment of p66 also plays a role in promoting dimerization. On the other hand, the presence of two 15-kDa C-terminal segments in the p66 homodimer is less favorable for dimerization than the presence of one 15-kDa C-terminal segment per dimer, by the quite significant free energy change of −1.25 kcal mol$^{-1}$. We and others (Barat et al., 1989) have noted within the central 20-kDa region of p66 a 29-residue segment (residues 283–310) of Leu repeats. This motif is highly conserved in HIV RTs and appears to have the potential to be a hepta-repeat leucine-zipper-like α-helix (Leu every seven residues) (Baillon et al., 1991). In addition, a 17-residue segment (residues 398–414) of Trp repeats (Trp every three or four residues) also is present within this 20-kDa central region. These two repeat motifs, containing a number of hydrophobic residues, could have implications for protein–protein binding.

The fact that RT is present only in low amounts in virions or in HIV-infected cells (Dickson et al., 1985) makes recombinant expression vectors a safer, more abundant, and far more flexible source of enzyme for biochemical and biophysical studies than HIV itself. It has been found that RT produced in bacteria is functionally active and shares many detailed enzymatic characteristics with the viral-derived enzyme (Majumdar et al., 1989; Lowe et al., 1988; Bebenek et al., 1989; Mizrahi, 1989). A complicating feature for studies of HIV RT is that the native enzyme appears to be a heterodimer of a larger and a smaller peptide, as discussed above. The origin or mechanism of formation of the smaller (i.e., the 51-kDa) peptide is not fully understood, although observations by Mizrahi et al. (1989) on the bacterial coexpression of HIV-1 protease and RT have suggested that p51 is derived from controlled proteolysis of p66 in situ by HIV-1 protease. From the standpoint of crystallization, domain mapping, and other detailed physical and enzymatic studies of RT, there are advantages to the use of preparations with just one polypeptide species, and a number of recombinant expression systems have been obtained for the 66-kDa HIV-1 RT (Tanese et al., 1986; Le Grice et al., 1987; Barr et al., 1987; Larder et al., 1987; Hansen et al., 1987; Hizi et al., 1988; Flexner et al., 1988; Farmerie et al., 1987; Müller et al., 1989). Most of these constructs have an ORF encoding one or more non-wild-type amino acid residues, as is the case with the constructs described here for p51, p29, and p15. The p66 constructs that are most similar to ours are by Hizi et al. (1988) and Larder et al. (1987) with respect to having a more precise RT sequence; these constructs were derived from HIV-1 strains different from our construct, LAV and BH10 proviral DNAs, respectively. Except for the amino-terminal Met, these constructs contain one non-wild-type residue at the amino terminus in the case of Hizi et al. (1988) and two non-wild-type amino-terminal residues in the case of Larder et al. (1987). The

functional significance, if any, of these extra residues in the amino terminus of the coding region is unknown. Finally, we note that a vector for production of a p51/p66 mixture (presumably heterodimer) corresponding to the precise RT coding region of BH10 DNA has been constructed (Mizrahi et al., 1989).

ACKNOWLEDGMENTS

We thank Dallas Lea for technical assistance, Judy Regan for oligonucleotide synthesis, and Dr. Bruno Starcich for providing pHX2D. We are grateful to Drs. J. Seehra and J. McCoy of Genetics Institute for helpful discussions and for providing RT used as a reference enzyme, to Dr. Madhuri Jaju for assistance with DNA polymerase enzymatic studies, to Drs. Jose Casas-Finet and Kenneth R. Williams for assistance with extinction coefficient measurements, and to Dr. Claude Klee for helpful discussions.

**Registry No.** RT, 9068-38-6.

REFERENCES

Abbotts, J., SenGupta, D. N., Zon, G., & Wilson, S. H. (1988) *J. Biol. Chem. 263*, 15094–15103.

Baillon, J. G., Nashed, N. T., Kumar, A., Wilson, S. H., & Jerina, D. M. (1991) *New Biologist 3*, 1015–1019.

Barat, C., Lullien, V., Schatz, O., Keith, G., Nugeye, M. T., Grüninger-Leitch, F., Barré-Sinoussl, F., LeGrice, S. F. J., & Darlix, J. C. (1989) *EMBO J. 8*, 3279–3285.

Barr, P. J., Power, M. D., Lee-Ng, C. T., Gibson, H. L., & Luciw, P. A. (1987) *Bio/Technology 5*, 486–489.

Bebenek, K., Abbotts, J., Roberts, J., Wilson, S. H., & Kunkel, T. A. (1989) *J. Biol. Chem. 264*, 16948–16956.

Becerra, P., Clore, G. M., Gronenborn, A. M., Karlstrom, A. R., Stahl, S. J., Wilson, S. H., & Wingfield, P. (1990) *FEBS Lett. 270*, 76–80.

Becerra, S. P., Koczot, F., Fabisch, P., & Rose, J. A. (1988) *J. Virol. 62*, 2745–2754.

Bernard, H. U., & Helinski, D. R. (1979) *Methods Enzymol. 68*, 482–492.

Chandra, A., Gerber, T., Kaul, S., Wolf, C., Demirhan, I., & Chandra, P. (1986) *FEBS Lett. 200*, 327–332.

Crowl, R., Seamans, C., Lomedico, P., & McAndrew, S. (1985) *Gene 38*, 31–38.

Dickson, C., Eisenman, R., Fan, H., Hunter, E., & Teich, N. (1985) in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses* (Weiss, R. A., Teich, N. M., Varmus, H. E., & Coffin, J. M., Eds.) pp 513–648, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Di Marzo-Veronese, F., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C., & Sarngadharan, M. G. (1986) *Science 231*, 1289–1291.

Farmerie, W. G., Loeb, D. D., Casavant, C., Hutchison, C. A., Edgell, M. H., & Swanstrom, R. (1987) *Science 236*, 305–308.

Ferris, A. L., Hizi, A., Showalter, S. D., Pichuantes, S., Babe, L., Craik, C. S., & Hughes, S. H. (1990) *Virology 175*, 456–464.

Flexner, C., Broyles, S. S., Earl, P., Chakrabarti, S., & Moss, B. (1988) *Virology 166*, 339–349.

Hanahan, D. J. (1983) *J. Mol. Biol. 166*, 557–580.

Hansen, J., Schulze, T., & Moelling, K. (1987) *J. Biol. Chem. 262*, 12393–12396.

Hansen, J., Schulze, T., Mellert, W., & Moelling, K. (1988) *EMBO J. 7*, 239–243.

Hizi, A., McGill, C., & Hughes, S. H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1218–1222.

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988) *Nature 331*, 280–283.

Karawya, E., Swack, J., & Wilson, S. H. (1983) *Anal. Biochem. 135*, 318–325.

Knott, G. D. (1979) *Comput. Programs Biomed. 10*, 271–280.

Kramer, R. A., Schaber, M. D., Skalka, A. M., Ganguly, K., Wong-Staal, F., & Reddy, E. P. (1986) *Science 231*, 1580–1584.

Kumar, A., Abbotts, J., Karawya, E. M., & Wilson, S. H. (1990) *Biochemistry 29*, 7156–7159.

Laemmli, U. K. (1970) *Nature 227*, 680–685.

Larder, B., Purifoy, D., Powell, K., & Darby, G. (1987) *EMBO J. 6*, 3133–3137.

Le Grice, S. F. J., Beuck, V., & Mous, J. (1987) *Gene 55*, 95–103.

Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A., & Venkatesan, S. (1986) *J. Virol. 60*, 771–775.

Lowe, D. M., Aitken, A., Bradley, C., Darby, G. K., Larder, B. A., Powell, K. L., Purifoy, D. J. M., Tisdale, M., & Stammers, D. K. (1988) *Biochemistry 27*, 8884–8889.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265–275.

Majumdar, C., Stein, C., Cohen, J., Broder, S., & Wilson, S. H. (1989) *Biochemistry 28*, 1340–1346.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Mizrahi, V., Lazarus, G. M., Miles, L. M., Meyers, C. A., & Debouck, C. (1989) *Arch. Biochem. Biophys. 273*, 347–358.

Müller, B., Restle, T., Weiss, S., Gautel, M., Sczakiel, G., & Goody, R. S. (1989) *J. Biol. Chem. 264*, 13975–13978.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Dovan, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., & Wong-Staal, F. (1985) *Nature 313*, 277–284.

Restle, T., Müller, B., & Goody, R. S. (1990) *J. Biol. Chem. 265*, 8986–8988.

Roark, D. E. (1976) *Biophys. Chem. 5*, 185–196.

Starnes, M. C., & Cheng, Y.-C. (1989) *J. Biol. Chem. 264*, 7073–7077.

Swack, J. A., Karawya, E., Albert, W., Fedorko, J., Minna, J. D., & Wilson, S. H. (1985) *Anal. Biochem. 147*, 10–21.

Tanese, N., Sodroski, J., Haseltine, W. A., & Goff, S. P. (1986) *J. Virol. 59*, 743–745.

Towbin, J., Staehelin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350–4354.

## APPENDIX: ULTRACENTRIFUGAL ANALYSIS OF A MIXED ASSOCIATION

The procedure for the analysis of a mixed association wherein species A is in a reversible heterogeneous reaction with species B, which is also involved in a reversible homogeneous monomer–dimer reaction, involves two distinct steps. The first of these involves the analysis of the ultracentrifugal behavior of species A and B separately, including the determination of the equilibrium constant for the homogeneous monomer–dimer reaction of species B. The second step involves the determination of the equilibrium constant for the reaction of species A with species B. In this discussion, species A represents p51 and species B represents p66.

Since species A does not undergo homogeneous association, its concentration distribution as a function of radial position at ultracentrifugal equilibrium is described by

$$c_{r,A} = c_{b,A} \exp(A_A M_A (r^2 - r_b^2)) + \epsilon = c_{b,A} \exp(A_A M_A \delta r^2) + \epsilon \quad \text{(A1)}$$

The concentrations may be expressed on a gram per liter, mole per liter, or UV absorbance concentration scale. The concentration of species $A$ at the radial position of the cell bottom is denoted by $c_{b,A}$; $A_A = (1 - \bar{v}_A \rho)\omega^2/2RT$, where $\bar{v}_A$ is the partial specific volume of species A, $\rho$ is the solution density, $\omega$ is the rotor angular velocity in radians per second, $R$ is the gas constant, and $T$ is the absolute temperature; $M_A$ is the molecular mass of species A; the base-line error is denoted by $\epsilon$. When using the UV absorption optical system, as was done here, the concentration terms used are absorbances, usually at a wavelength of 280 nm. Since the molecular mass of species A is known exactly from amino acid sequence data, the purpose of this analysis is to demonstrate the ideal monomeric behavior of the species, to determine experimentally the value of $A_A$ and thus also the value of $\bar{v}_A$, and to experimentally determine the initial concentration, $c_{0,A}$, as will be discussed later. The terms $c_{b,A}$, $A_A$, and $\epsilon$ are fitting parameters which are determined in the nonlinear least-squares curve-fitting procedure.

The observed concentration distribution of species B as a function of radial position at ultracentrifugal and chemical equilibrium is described by

$$c_{r,B} = c_{b,B} \exp(A_B M_B \delta r^2) + k_{12} c_{b,B}^2 \exp(2A_B M_B \delta r^2) + \epsilon \quad \text{(A2)}$$

where the equilibrium constant for dimer formation is defined as $k_{12} = c_{b,B_2}/c_{b,B}^2$; the concentrations are expressed as absorbances here. The molar equilibrium constant, defined as $K_{12} = C_{b,B_2}/C_{b,B}^2$, with the concentrations expressed as moles per liter, may be calculated from this observed equilibrium constant. The molar concentration is related to the absorbance by $C = c/E$, where $E$ is the molar extinction coefficient. Making the reasonable assumption that $E_2 = 2E_1$, then $K_{12} = k_{12}E_1/2$.

Equation A2 implicitly assumes that the dimer has the same partial specific volume as the monomer. This may be verified by studying the association at different rotor speeds or different solution column lengths which will produce different hydrostatic pressures within the cell. If the monomer and dimer have different partial specific volumes, then the apparent association constant will depend upon the hydrostatic pressure, and if the difference is of sufficient magnitude, it may be detectable in such experiments.

Using eq A2 as a mathematical model for fitting the data in order to obtain the value of the equilibrium constant is not optimal. For several reasons it is better to use the equation

$$c_{r,B} = c_{b,B} \exp(A_B M_B \delta r^2) + \exp(\ln k_{12}) c_{b,B}^2 \exp(2A_B M_B \delta r^2) + \epsilon \quad \text{(A3)}$$

This mathematical model assures that $k_{12}$ is physically meaningful by virtue of having a positive value. The use of $\ln k_{12}$ thus acts as an implicit constraint built into the mathematical model. Fitting for $\ln k_{12}$ also has the advantage that the quantity $\ln K_{12}$, given by $\ln K_{12} = \ln k_{12} + \ln (E_1/2)$, is used to calculate thermodynamic parameters such as $\Delta G°$, and the fitting procedure will directly give the estimated error in this parameter. Finally, the curve describing the calculated values of the sum of squares obtained as a function of values of the parameter $k_{12}$ is asymmetrical about the optimal value of the parameter, indicating that the error estimates are poor. However, if $\ln k_{12}$ is used, the sum of squares curve is very nearly a parabola, symmetrical about the optimal value of $\ln k_{12}$, indicating that the error estimates are more nearly valid. This test is based on the fact that for a linear model where the data has normally distributed error and where the data is properly weighted in the fitting procedure, a graph of the sum of squares as a function of the values of a parameter will be a parabola that is symmetrical about the optimal parameter value. This becomes a heuristic test for the nonlinear case. In addition to obtaining the value of $\ln k_{12}$, fitting with eq A3 is used to obtain an optimal value of $\bar{v}_B$. Initially, the data is fit using a value of $A_B$ incorporating a value of $\bar{v}_B$ calculated from the amino acid composition. Then, the data is fit again permitting the value of $A_B$ to be varied. If the root mean square error and the distribution of the residuals are improved, then the value obtained by fitting is used; if this is not the case, then the compositional value is used. The initial concentration, $c_{0,B}$, can be calculated from this as will be discussed later.

When the value of $\ln k_{12}$ and the values of the partial specific volumes have been obtained, then it is possible to fit the data obtained when a mixture of the two components is centrifuged. The total concentration as a function of radial position when such a mixture is at ultracentrifugal and chemical equilibrium is given by

$$c_r = c_{b,A} \exp(A_A M_A \delta_r^2) + c_{b,B} \exp(A_B M_B \delta r^2) + \exp(\ln k_{12}) c_{b,B}^2 \exp(2A_B M_B \delta r^2) + \exp(\ln k_{AB}) c_{b,A} c_{b,B} \exp(A_{AB} M_{AB} \delta r^2) + \epsilon \quad \text{(A4)}$$

This equation assumes that the partial specific volume of the AB complex is a weight average partial specific volume which may be calculated from the partial specific volumes of the species forming it. Thus, $\bar{v}_{AB} = (M_A \bar{v}_A + M_B \bar{v}_B)/M_{AB}$. While it should be possible to validate this assumption by dilatometric experiments, it is usually accepted without verification because of the difficulty of the experiments and the quantities of the reactants required.

Attempts to fit mixed association data with eq A4 as a mathematical model will frequently fail, and did so in this study. It appears that attempting to fit this model is an ill-conditioned problem since the surface defined by the values of the sum of squares as a function of the parameter values has a broad, shallow, and ill-defined minimum. Convergence is slow, and even with a very stringent convergence limit, the parameter values obtained frequently have values that are physically unrealistic even when the sum of squares has an apparently acceptable minimum value and the distribution of the residuals also appears to be acceptable.

The solution to this problem lies in the application of the method of implicit constraints (Lewis & Youle, 1986; Lewis, 1987; Lewis, et al., 1988; Bubb et al., 1991). This method develops a fitting model which takes into account conservation of mass of the components in the centrifuge cell and relates the concentrations of the components at equilibrium to their initial loading concentrations and to each other. In doing this it eliminates one concentration value as a parameter and, more important, changes the shape of the sum of squares surface so that it appears to have a relatively deep and well-defined minimum. As a result, convergence is rapid and the parameter values are realistic.

The total concentrations of species A and B as a function of radial position, expressed as moles of monomer per liter and denoted as $C_{r,A,T}$ and $C_{r,B,T}$, are given by

$$C_{r,A,T} = C_{b,A} \exp(A_A M_A \delta r^2) + C_{b,AB} \exp(A_{AB} M_{AB} \delta r^2) \quad (A5)$$

$$C_{r,B,T} = C_{b,B} \exp(A_B M_B \delta r^2) + 2C_{b,B_2} \exp(A_{B_2} M_{B_2} \delta r^2) + C_{b,AB} \exp(A_{AB} M_{AB} \delta r^2) \quad (A6)$$

The effective initial loading concentration of each component may be obtained from conservation of mass considerations. Conservation of mass in the ultracentrifuge cell is described by

$$C_0(r_b^2 - r_m^2) = C_0 \Delta r^2 = \int_{r_m}^{r_b} C_r \, dr^2 \quad (A7)$$

where $C_0$ is the initial concentration of the species and $r_m$ and $r_b$ are the radial positions of the solution meniscus and the cell bottom, respectively. Since the concentration distribution of any species at equilibrium is described by

$$C_r = C_b \exp(AM\delta r^2) \quad (A8)$$

the initial concentration of that species is given by substituting eq A8 in eq A7, performing the integration and rearranging to obtain

$$C_0 = (C_b/AM)(1 - \exp(AM\Delta r^2)) \quad (A9)$$

This equation is valid only for the conventional centerpiece whose sectoral walls lie on radii from the center of rotation of the rotor. While eqs A7, A8, and A9 have been written using the mole per liter concentration scale, they are equally valid for the concentration expressed in any other units.

The total concentrations of species A and B can be obtained by performing the appropriate integrations of eqs A5 and A6. Performing these integrations using the simplifying conventions that $\Phi_A = A_A M_A \Delta r^2$, $\Phi_B = A_B M_B \Delta r^2$, $\Phi_{AB} = A_{AB} M_{AB} \Delta r^2$, and $\Phi_{B_2} = A_{B_2} M_{B_2} \Delta r^2 = 2\Phi_B$ then gives the total concentrations of these species, expressed as moles of monomer and given by

$$C_{0,A} = (C_{b,A}/\Phi_A)(1 - \exp(-\Phi_A)) + (C_{b,AB}/\Phi_{AB})(1 - \exp(-\Phi_{AB})) \quad (A10)$$

$$C_{0,B} = (C_{b,B}/\Phi_B)(1 - \exp(-\Phi_B)) + (2C_{b,B_2}/2\Phi_B) \times (1 - \exp(-2\Phi_B)) + (C_{b,AB}/\Phi_{AB})(1 - \exp(-\Phi_{AB})) \quad (A11)$$

It should be noted that the final terms of eqs A10 and A11 are identical. These equations can then be rearranged and combined to give

$$C_{0,A} - (C_{b,A}/\Phi_A)(1 - \exp(-\Phi_A)) = C_{0,B} - (C_{b,B}/\Phi_B)(1 - \exp(-\Phi_B)) - (2C_{b,B_2}/2\Phi_B)(1 - \exp(-2\Phi_B)) \quad (A12)$$

which can then be solved for $C_{b,A}$. This is simplified by defining the following calculatable terms

$$\psi_A = (1 - \exp(-\Phi_A))/\Phi_A \quad (A13)$$

$$\psi_B = (1 - \exp(-\Phi_B))/\Phi_B \quad (A14)$$

$$\psi_{2B} = (1 - \exp(-2\Phi_B))/2\Phi_B \quad (A15)$$

Converting to the absorbance concentration scale now gives

$$c_{0,A}/E_A - c_{b,A}\psi_A/E_A = c_{0,B}/E_B - c_{b,B}\psi_B/E_B - 2c_{b,B_2}\psi_{2B}/2E_B \quad (A16)$$

Solving now for $c_{b,A}$ and using $c_{b,B_2} = k_{12} c_{b,B}^2$ gives

$$c_{b,A} = (c_{0,A}E_B - c_{0,B}E_A + c_{b,B}\psi_B E_A + k_{12} c_{b,B}^2 \psi_{2B} E_A)/\psi_A E_B \quad (A17)$$

Further simplification may be effected by defining the following calculatable terms:

$$D_{AB} = (c_{0,A}E_B - c_{0,B}E_A)/\psi_A E_B \quad (A18)$$

$$R_1 = \psi_B E_A/\psi_A E_B \quad (A19)$$

$$R_2 = \psi_{2B} E_A/\psi_A E_B \quad (A20)$$

Substituting these in eq A17 then gives

$$c_{b,A} = D_{AB} + R_1 c_{b,B} + R_2 \exp(\ln k_{12}) c_{b,B}^2 \quad (A21)$$

This can now be substituted in eq A4, which, with appropriate factoring, gives the constrained model

$$c_r = [D_{AB} + R_1 c_{b,B} + R_2 \exp(\ln k_{12}) c_{b,B}^2] \times [\exp(A_A M_A \delta r^2) + (\exp(\ln k_{AB}) c_{b,B} \times \exp(A_{AB} M_{AB} \delta r^2))] + c_{b,B}[\exp(A_B M_B \delta r^2) + (\exp(\ln k_{12}) c_{b,B}^2 \exp(2A_B M_B \delta r^2))] + \epsilon \quad (A22)$$

In this model, only $\ln k_{AB}$, $c_{b,B}$, and $\epsilon$ are fitting parameters; all of the other terms are known or have been determined from the previous experiments on the individual components.

It is apparent that the value of the term $D_{AB}$ will be zero if the loaded reactant concentrations are equimolar. If this condition is met, or if the concentrations are nearly equimolar, then small errors in the concentrations will have minimal effect on the results. However, it is still desirable to verify that mass has been conserved in the centrifuge cell. The optimal method for achieving this is to use information from the experiments for the characterization of the individual components. The effective values of $c_0$ for each of these components can be obtained by the appropriate integration of eqs A1 and A3 which give, respectively

$$c_{0,A} = (c_{b,A}/\Phi_A)(1 - \exp(-\Phi_A)) \quad (A23)$$

$$c_{0,B} = (c_{b,B}/\Phi_B)(1 - \exp(-\Phi_B)) + (\exp(\ln k_{12}) c_{b,B}^2)(1 - \exp(-2\Phi_B)) \quad (A24)$$

and where the values of the various terms are appropriate to the specific experiment. If these same solutions of the individual components are used to make up the solution which is analyzed for the interaction, then the values of $c_0$ of each component in the mixture can be calculated and used in eq A22. Accurate knowledge of these values requires either accurate pipetting or, better, accurate weighing of the quantity of each solution used.

It is not possible to fit the mixed association concentration distribution with a different model which is analogous to eq A4 with the expectation of determining the individual $c_0$ values since the nature of the sum of squares surface as a function of the parameter values is so ill-defined as to preclude this. However, it appears to be possible to obtain the sum of the individual $c_0$ values if the model is formulated appropriately. The model to be used, analogous to eq A4, is

$$c_r = \exp(\ln c_{b,A}) \exp(A_A M_A \delta r^2) + \exp(\ln c_{b,B}) \exp(A_B M_B \delta r^2) + \exp(\ln c_{b,B_2}) \exp(2A_B M_B \delta r^2) + \exp(\ln c_{b,AB}) \exp(A_{AB} M_{AB} \delta r^2) + \epsilon \quad (A25)$$

The logarithms of the concentrations are used as fitting parameters in order to avoid the possibility of negative values for the concentrations. The total value of $c_0$ is then given by

$$c_{0,T} = (\exp(\ln c_{b,A})/\Phi_A)(1 - \exp(-\Phi_A)) + (\exp(\ln c_{b,B})/\Phi_B)(1 - \exp(-\Phi_B)) + (\exp(\ln c_{b,B_2})/2\Phi_B)(1 - \exp(-2\Phi_B)) + (\exp(\ln c_{b,AB})/\Phi_{AB})(1 - \exp(-\Phi_{AB})) \quad (A26)$$

This value should be equal to the sum of the individual $c_0$

values used in eqs A18 and A22.

REFERENCES

Bubb, M. R., Lewis, M. S., & Korn, E. D. (1991) *J. Biol. Chem. 266*, 3820–3826.

Lewis, M. S. (1987) *Biophys. J. 51*, 441a.

Lewis, M. S., & Youle, R. J. (1986) *J. Biol. Chem. 261*, 11571–11577.

Lewis, M. S., Luborsky, S. W., & Yamada, K. M. (1988) *Biophys. J. 53*, 75a.

# DNA Cross-Linking and Sequence Selectivity of Aziridinylbenzoquinones: A Unique Reaction at 5′-GC-3′ Sequences with 2,5-Diaziridinyl-1,4-benzoquinone upon Reduction†

John A. Hartley,*,‡ Mark Berardini,‡ Mauro Ponti,‡ Neil W. Gibson,§ Andrew S. Thompson,∥ David E. Thurston,∥ Brigid M. Hoey,⊥ and John Butler⊥

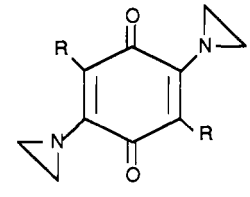
*Department of Oncology, University College and Middlesex School of Medicine, 91 Riding House Street, London W1P 8BT, U.K., Department of Pharmaceutical Sciences, School of Pharmacy, and Comprehensive Cancer Center, University of Southern California, Los Angeles, California 90033, Division of Medicinal Chemistry, School of Pharmacy and Biomedical Sciences, Portsmouth Polytechnic, Park Building, King Henry 1st Street, Portsmouth, Hampshire PO1 2DZ, U.K., and CRC Department of Biophysical Chemistry, Paterson Institute for Cancer Research, Christie Hospital and Holt Radium Institute, Wilmslow Road, Manchester M20 9BX, U.K.*

*Received July 10, 1991; Revised Manuscript Received September 17, 1991*

ABSTRACT: Several bifunctional alkylating agents of the aziridinylbenzoquinone class have been evaluated as potential antitumor agents. 3,6-Bis[(2-hydroxyethyl)amino]-2,5-diaziridinyl-1,4-benzoquinone (BZQ), 2,5-diaziridinyl-1,4-benzoquinone (DZQ), 3,6-bis(carboxyamino)-2,5-diaziridinyl-1,4-benzoquinone (AZQ), and six analogues of AZQ have been studied for their ability to induce DNA interstrand cross-linking, as measured by an agarose gel technique, and to determine whether they react with DNA in a sequence-selective manner, as determined by a modified DNA sequencing technique. At an equimolar concentration (10 μM), only DZQ and BZQ showed any detectable cross-linking at pH 7 without reduction. Cross-linking was enhanced in both cases at low pH (4). Reduction by ascorbic acid at both pH's increased the cross-linking, which was particularly striking in the case of DZQ. In contrast, AZQ and its analogues only produced a significant level of cross-linking under both low-pH and reducing conditions, the extent of cross-linking decreasing as the size of the alkyl end group increased. The compounds reacted with all guanine-N7 positions in DNA with a sequence selectivity similar to other chemotherapeutic alkylating agents, such as the nitrogen mustards, although some small differences were observed with BZQ. Nonreduced DZQ showed a qualitatively similar pattern of reactivity to the other compounds, but on reduction (at pH 4 or 7) was found to react almost exclusively with 5′-GC-3′ sequences, and in particular, at 5′-TGC-3′ sites. A model to explain this unique reaction is proposed.

Several natural and synthetic quinones including adriamycin, mitomycin C, and aziridinylbenzoquinones have found an application as antitumor agents. Such compounds have the potential to undergo reduction by cellular enzymes to produce more active forms (Butler & Hoey, 1987; Powis, 1987) although it has been difficult to prove the role of bioreductive activation in the antitumor activity or the production of toxic side effects of these quinones, and a single mechanism cannot fully explain all the observed cytotoxic effects.

Two aziridinylbenzoquinones, AZQ (diaziquone, Figure 1, compound D2) and BZQ (Figure 1), have undergone clinical trials as potential antitumor drugs (Khan & Driscoll, 1979; Bender et al., 1983; Haid et al., 1985). Aziridines have the potential to react with available nucleophiles to form ring-opened covalent adducts. Many aziridinylbenzoquinones have the added advantage in that the ring-opening process would



| COMPOUND | R |
|---|---|
| BZQ | $NHCH_2CH_2OH$ |
| DZQ | H |
| D1 | $NHCOOCH_3$ |
| D2 (AZQ) | $NHCOOC_2H_5$ |
| D3 | $NHCOOC_3H_7$ (*n*-propyl) |
| D4 | $NHCOOC_3H_7$ (*i*-propyl) |
| D5 | $NHCOOC_4H_9$ (*n*-butyl) |
| D6 | $NHCOOC_4H_9$ (*i*-butyl) |
| D7 | $NHCOOC_4H_9$ (*sec*-butyl) |

FIGURE 1: Structures of the aziridinylbenzoquinones used in this study.

† This work was funded in part by the Cancer Research Campaign and by a collaborative NATO grant (0143/89) to J.A.H. and N.W.G. M.B. thanks Dixons for a predoctoral studentship.

* To whom correspondence should be addressed.

‡ University College and Middlesex School of Medicine.

§ University of Southern California.

∥ Portsmouth Polytechnic.

⊥ Christie Hospital and Holt Radium Institute.

0006-2960/91/0430-11719$02.50/0 